BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
26 April 2001 (26.04.2001)

PCT

(10) International Publication Number
WO 01/29212 A1

(51) International Patent Classification[7]: C12N 15/10, C12Q 1/68, C12P 19/34

(21) International Application Number: PCT/US00/41319

(22) International Filing Date: 19 October 2000 (19.10.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/160,420	19 October 1999 (19.10.1999)	US
09/514,660	29 February 2000 (29.02.2000)	US
60/219,062	18 July 2000 (18.07.2000)	US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US	09/514,660 (CIP)
Filed on	29 February 2000 (29.02.2000)
US	60/219,062 (CIP)
Filed on	18 July 2000 (18.07.2000)

(71) Applicant *(for all designated States except US)*: ENCHIRA BIOTECHNOLOGY CORPORATION [US/US]; 4200 Research Forest Drive, The Woodlands, TX 77381 (US).

(72) Inventors; and

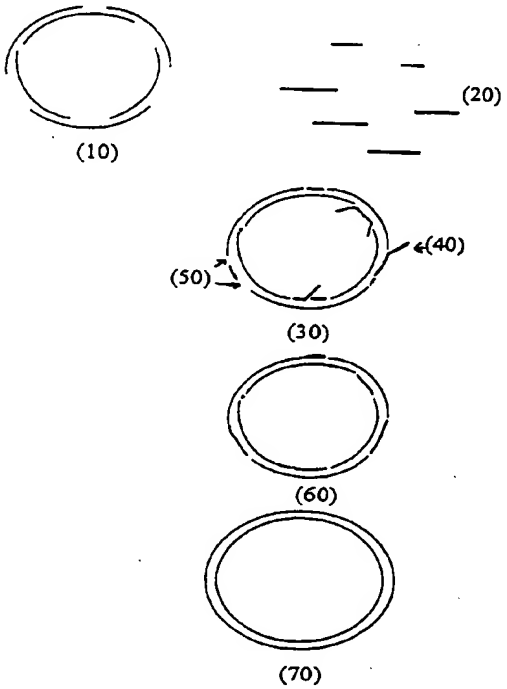
(75) Inventors/Applicants *(for US only)*: ARENSDORF, Joseph, J. [US/US]; 27225 Orth Lane, Oak Ridge, TX 77385 (US). COCO, Wayne, M. [US/US]; 23 Cornerbrook Place, The Woodlands, TX 77381 (US).

(74) Agents: ELMORE, Carolyn, S. et al.; Hamilton, Brook, Smith & Reynolds, P.C., Two Militia Drive, Lexington, MA 02421 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR,

[Continued on next page]

(54) Title: METHODS FOR CHIMERAGENESIS OF WHOLE GENOMES OR LARGE POLYNUCLEOTIDES



(57) Abstract: A method for forming at least one contiguous double-stranded chimeric polynucleotide comprising modifying a double-stranded polynucleotide of interest such that single-stranded nicks are introduced on at least one region of one strand of the polynucleotide, treating the nicked polynucleotide with a nuclease such that at least one single-stranded region of the polynucleotide are, contacting said polynucleotide comprising the single-stranded region with a random population of oligonucleotides such that at least one oligonucleotide hybridizes to at least one single-stranded region, forming a double-stranded hybridized complex; and treating the hybridized complex such that at least one contiguous chimeric polynucleotide is formed is provided.

WO 01/29212 A1

HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) **Designated States** *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— *With international search report.*

— *Before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

METHODS FOR CHIMERAGENESIS OF WHOLE GENOMES OR LARGE POLYNUCLEOTIDES

RELATED APPLICATIONS

This application is a Continuation-in-Part of United States Application No: 09/514,660, filed February 29, 2000 which claims the benefit of United States Provisional Application No. 60/160,420, filed October 19, 1999. This application also claims the benefit of United States Provisional Application No. 60/219,062, filed July 18, 2000. The teachings of these applications are incorporated by reference herein in their entireties.

BACKGROUND OF THE INVENTION

Genetic improvements occur more frequently when the generation of mutations is coupled with genetic recombination. The effect of genetic recombination on the fixing in a population of multiple beneficial mutations is evident when comparing sexually versus asexually replicating organisms. Asexually replicating organisms exhibit an accumulation of mutations that limit their evolutionary potential because they have no mechanism to combine beneficial mutations while eliminating detrimental mutations due to the fact that genetic recombination does not occur. This reduction in evolutionary potential in asexually replicating populations is known as Müller's ratchet (Müller, H., *Mut. Res.* 1964. 1:2-9). Recombination between altered or otherwise non-identical polynucleotide targets allows the consolidation of favorable mutations that originally occurred on separate copies of the target, as well as the elimination of deleterious mutations (Harayama, S., *Trends Biotechnol.* 1998. 16:76-82). The generation of genetic

diversity (mutagenesis), including recombination, coupled with selection and screening, is termed "directed evolution."

Methods of directed evolution include the "sexual PCR" method, which involves cleaving a population of target DNA using DNase I followed by the reassembly of cleaved DNA fragments. Reassembly of DNA fragments results in partially double-stranded DNA molecules which are then used as templates during a self-primed Polymerase Chain Reaction (hereinafter "PCR") step (Stemmer, W. P., *Nature*. 1994. 370:389-9; U.S. Patent Nos. 5,605,793 and 5,811,238).

Alternative methods for directed evolution involve a modification of the sexual PCR method known as "Random Priming Recombination" (hereinafter "RPR"; Shao, Z. *et al.*, *Nucleic Acids Res.* 1998. 26:681-3), as well as alternative methods such as the "Staggered Extension Process" (hereinafter "StEP"; Zhao, H. *et al.*, *Nat. Biotechnol.* 1998. 16:258-61). An essential step for both of these methods is a PCR step. RPR utilizes elongation of randomly annealed primers to synthesize fragments from a given template. These fragments are then used for the reassembly step in the sexual PCR method. StEP involves the binding of a primer to the terminus of a mutant or wild-type template. A brief round of polymerization extends the primer, however, extension is prematurely halted, resulting in a DNA fragment shorter than the original template. The short fragment is denatured from its original template and allowed to reanneal, potentially to a template with a different mutation. Subsequent rounds of short polymerization and denaturation further extend the fragment, eventually producing a newly synthesized strand that is as long as the template strand.

Another method of directed evolution is described in WO 00/09679 in which various types of assembly matricies are utilized. Typically, repeated rounds of annealing, ligation and denaturation are required to generate target-length chimeric polynucleotides using the method.

The result of these methods of directed evolution are "chimeric" polynucleotides, so called because they include sequences from more than one parent template. These methods all require multiple steps to generate target length chimeric molecules. Sexual PCR, RPR and StEP each result in a library of

recombinant DNA chimeras of the starting parental genes, however, due to limitations inherent in the essential PCR step in each of these methods, the library of chimeric products generally represent a limited sampling of all potential chimeric products. Furthermore, use of large nucleic acid molecules as template, or nucleic acid comprising more than one gene is often problematic in these methods and result in low rates of mosaicism or high numbers of inactive chimeras within the chimeric library.

SUMMARY OF THE INVENTION

The methods of the present invention facilitate the generation of chimeric polynucleotides using templates comprising one or more single-stranded regions. The method is particularly useful where the template comprises a large nucleic acid molecule. A double-stranded polynucleotide containing at least one single-stranded region is contacted with a population of oligonucleotides. The single-stranded regions serve to precisely order and align hybridizing oligonucleotides. The oligonucleotides are allowed to hybridize to the single-stranded regions of the template and recombine to form double-stranded chimeric polynucleotides. The hybridized oligonucleotides are ligated to form a contiguous chimeric heteroduplex polynucleotide. The final chimeric product is a heteroduplex that typically does not contain a full-length template uninterrupted by hybridized oligonucleotides.

The method is particularly suited for generating libraries of chimeric double-stranded polynucleotides containing genotypic variants that have desired properties (phenotype) that are improved with respect to parental phenotypes. The improvements can be directed to individual polypeptides, to entire pathways, to cellular regulatory systems, or to cellular structures, *e.g.*, cell wall. The invention is applicable to prokaryotic systems as well as to eukaryotic systems such as yeast, plant and even human systems.

In one embodiment, the present invention is directed to a method for forming a double-stranded chimeric polynucleotide including the steps of contacting a double-stranded polynucleotide template containing at least two single-stranded regions with a population of oligonucleotides such that at least one oligonucleotide

hybridizes to each single-stranded region, and treating the template and hybridized oligonucleotides, thereby forming a double-stranded chimeric polynucleotide. The template and hybridized oligonucleotides can be treated, for example, with a ligase, such that adjacently hybridized oligonucleotides are ligated, forming a double-stranded chimeric polynucleotide.

In another embodiment, the method for forming a double-stranded polynucleotide also includes the step of preparing a polynucleotide template containing at least one single-stranded region. Preparation of the template includes modifying a double-stranded polynucleotide such that at least one single-stranded nick is introduced and treating the nicked double-stranded polynucleotide such that a single-stranded region is formed at each nick. In a particular embodiment, the double-stranded polynucleotide is modified by treating the double-stranded polynucleotide with DNase I. The nicked double-stranded polynucleotide can be treated, for example, with a nuclease, such as Exonuclease III, to form single-stranded regions.

In other embodiments, the method can include the steps of trimming the flaps of hybridized oligonucleotides or filling in gaps that may be present between hybridized oligonucleotides or between hybridized oligonucleotides and a template strand of the same orientation. The method of forming double-stranded chimeric molecules can also include combinations of trimming and gap filling.

The method of forming double-stranded chimeric polynucleotides can further comprise a step of selecting or screening the double-stranded chimeric polynucleotide for a specified characteristic. The specified characteristic can be altered, for example, in comparison to a reference polynucleotide.

The invention is also drawn to a method for preparing a double-stranded polynucleotide template suitable for use in forming a double-stranded chimeric polynucleotide. The method of preparing a double-stranded template includes the steps of treating a double-stranded polynucleotide such that at least two single-stranded nicks are introduced and treating the nicked polynucleotide such that a single-stranded region of the polynucleotide is formed at each nick, thereby resulting in a double-stranded polynucleotide suitable for use in forming a chimeric

polynucleotide. The double-stranded polynucleotide can be, for example, isolated from a suitable nucleic acid source, synthetically manufactured or cleaved from a larger polynucleotide. The double-stranded polynucleotide can be amplified using methods well known in the art.

The present invention is also drawn to chimeric double-stranded polynucleotides produced by the method described herein.

The present invention allows chimeragenesis of genes or gene clusters that are not specifically localized but are known to be on a particular cosmid or genome. Shearing of longer templates can be minimized, for example, by conducting the steps of the method in a semi-solid medium, such as agarose.

BRIEF DESCRIPTION OF THE DRAWING

The foregoing and other objects, features and advantages of the invention will be apparent from the following more particular description of preferred embodiments of the invention, as illustrated in the accompanying drawing.

The Figure is a schematic diagram showing one embodiment of the method of the present invention.

DETAILED DESCRIPTION OF THE INVENTION

The present invention facilitates the generation of chimeric polynucleotides using double-stranded templates that comprise one or more single-stranded regions. "Chimeric polynucleotides," as used herein, typically contain nucleotide sequences from multiple related sequences or otherwise similar polynucleotides, referred to herein as "parent polynucleotides."

The method is particularly useful where the template comprises a large nucleic acid molecule, such as is common with genomic DNA, nucleic acid comprising more than one gene, including chromosomes or fragments thereof and operons, or nucleic acid encoding large molecules including polyketide synthetases or non-ribosomal peptide synthetases.

The methods described herein comprise process steps involved in the formation of chimeric polynucleotides. Reference is now made to the Figure. The

Figure depicts schematically the steps utilized by one embodiment of the present invention in forming the double-stranded chimeric polynucleotide, wherein a single-stranded region is present on each strand of the double-stranded template. A circular double-stranded polynucleotide is prepared such that it comprises at least one single-stranded region on each strand 10. A population of related oligonucleotides 20 is assembled onto the single-stranded regions of the template in such a way as to allow for oligonucleotides derived from different parent polynucleotides to assemble onto the same template and even within the same single-stranded region 30. In some cases, overlaps occur, thus creating "flaps" 40. The term "flaps" is intended to include the unhybridized portion of an oligonucleotide that is otherwise hybridized to a template. In other cases, regions of the template remain single-stranded, thus creating "gaps" 50. Flaps can be trimmed and gaps can be filled in prior to the generation of a contiguous chimeric polynucleotide 60. The contiguous double-stranded chimeric polynucleotide can be generated by ligating the assembled oligonucleotides. The product double-stranded chimeric polynucleotide can then be selected or screened for optimized traits of interest.

In one embodiment, the present invention is directed to a method for forming a double-stranded chimeric polynucleotide comprising the steps of contacting a double-stranded polynucleotide template containing at least one single-stranded region with a population of oligonucleotides, also referred to herein as donor fragments, such that at least one oligonucleotide hybridizes to each single-stranded region; and treating the template and hybridized oligonucleotides, thereby forming a double-stranded chimeric polynucleotide. The single-stranded region or regions, also referred to herein as target sequences, serve to precisely order and align hybridizing oligonucleotides. The hybridization can be conducted under conditions of low stringency. In a particular embodiment, the double-stranded polynucleotide template contains at least two nicks, resulting in two single-stranded regions. The double-stranded template polynucleotide can be from about 200 base pairs to about 20 megabase pairs in length and the single-stranded regions can be from about 5 bases to about 50,000 bases in length. Shearing of longer templates can be

minimized, for example, by conducting the steps of the method in a semi-solid medium, such as agarose.

In another embodiment, the method for forming double-stranded chimeric polynculeotides also includes the step of preparing a polynucleotide template containing at least two single-stranded regions. Preparation of the template includes modifying a double-stranded polynucleotide such that at least two single-stranded nicks are introduced; and treating the nicked double-stranded polynucleotide such that a single-stranded region is formed at each nick. In a particular embodiment, the nicks are introduced at random locations. The nicks can be introduced on either or both strands of the template. In a particular embodiment, the double-stranded polynucleotide is modified by treating the double-stranded polynucleotide with DNase I. Other methods for introducing single-stranded nicks or breaks such as 1 base gaps include incorporation of uracil into the double-stranded template, treating the template with UDP glycosylase followed by apurine endonuclease to generate one base gaps. In this method, uracil can be incorporated into one or both strands of the template. Other methods of introducing single-stranded nicks include methods describe in P. E. Nielsen, *J. Molecular Recognition 3:*1-25 (1990), including exposing the template to ionizing radiation, or treating the template with osmium tetroxide, potassium permanganate or ethylnitrosourea under suitable conditions to generate single-stranded nicks.

The nicked, double-stranded polynucleotide is then treated to create single-stranded regions. In a particular embodiment, the nicked double-stranded polynucleotide is treated with Exonuclease III (herein after "Exo III"). While not wishing to be bound by theory, Exo III is an exonuclease that produces single-stranded gaps in double-stranded DNA by catalyzing the step-wise removal of nucleotides from the 3'-terminus at each nick. The treated double-stranded polynucleotide template contains at least one region that is single-stranded. In other embodiments, single-stranded regions can be generated at the nicks by treating the nicked template with exonucleases that remove nucleotides from the 5'-terminus at each nick. These enzymes include lambda exonuclease and T_7 Gene 6 Exonuclease. In still another embodiment, the nicked template can be treated with exonucleases

that remove nucleotides from the 3' terminus and an exonuclease that remove nucleotides from the 5' terminus such that single-stranded regions are produced that extend in both directions from the nick.

It is clear that one of ordinary skill in the art can modify the conditions of the method to generate the desired density of nicks in the double-stranded template, the desired length of single-stranded regions or combinations thereof. For example, conditions can be chosen to introduce short single-stranded regions from a large number of nicks. Nicks can be introduced, for example, such that a nick is introduced at about every 500 nucleotides. In another embodiment, a nick is introduced at about every 250 nucleotides, in still another embodiment a nick is introduced at about every 200 nucleotides. Short single-stranded regions could be used, for example, to generate chimeric molecules involving particular regions of a template, such as regulatory regions or regions encoding domains or subdomains. In another example, conditions can be chosen to introduce long single-stranded regions from a small number of nicks. Nicks can be introduced, for example, such that a nick is present at about every 1000 nucleotides. In another embodiment, a nick is present at about every 5000 nucleotides, in still another embodiment, a nick is introduced at about every 10,000 nucleotides or at about every 50,000 nucleotides. Long single-stranded regions could be used, for example, to generate chimeric polynucleotides comprising large genes, templates comprising more than one gene, genomes, or fragments thereof or chromosomes or fragments thereof, such that whole genes or regions comprising multiple domains are exchanged or chimerized with donor fragments.

The single-stranded regions can be from about 25 nucleotides to about 50,000 nucleotides in length. In other embodiments, the single-stranded regions can be from about 250 to about 10,000 nucleotides in length; or from about 250 to about 5,000 nucleotides in length. The single-stranded regions can be at least about 25 nucleotides in length, at least about 250 nucleotides in length or at least about 500 nucleotides in length. In one embodiment, a single-stranded region comprise about 10% of the length of the double-stranded template.

In other embodiments, the method can include the steps of trimming flaps of hybridized oligonucleotides and/or filling in any gaps that may be present between hybridized oligonucleotides or hybridized oligonucleotides and template of the same orientation. Combinations of trimming and gap filling are also within the scope of the invention.

The method further comprises a step of selecting or screening a chimeric polynucleotide for a specified characteristic. The specified characteristic can be altered, for example, in comparison to a reference polynucleotide.

The invention is also drawn to a method for preparing a double-stranded polynucleotide template suitable for use in forming a double-stranded chimeric polynucleotide. The method of preparing a double-stranded template includes the steps of isolating a double-stranded polynucleotide; treating the double-stranded polynucleotide such that at least two single-stranded nicks are introduced; and treating the nicked polynucleotide such that a single-stranded region of the polynucleotide is formed at each nick, thereby resulting in a double-stranded polynucleotide suitable for use in forming a chimeric polynucleotide.

The nucleic acid for use as a template, parent polynucleotides or donor fragments can be synthetically manufactured or isolated from any suitable source of nucleic acid. The template, parent polynucleotides or donor fragments of the present invention can comprise DeoxyriboNucleic Acid (hereinafter "DNA"), or RiboNucleic Acid (hereinafter "RNA") DNA or RNA can comprise natural bases, *e.g.*, adenine, thymine, cytosine, guanine or uracil; analog bases, *e.g.*, inosine, bromouracil or nitroindole; chemically altered bases, *e.g.*, biotin labeled or digoxygenin labeled bases; or a combination thereof provided that the resulting chimeric polynucleotide can be replicated. The template, parent polynucleotides or donor fragments can be derived from a larger molecule, *e.g.*, can be a restriction fragment, or can be the product of an enzymatic amplification or chemical synthesis *in vitro*.

The double-stranded template of the present invention can be from about 200 bases to about 20 megabases in length. In particular embodiments, the double-stranded template can be from about 100 kilobases to about 10 megabases in length,

from about 100 kilobases to about 350 kilobases in length or from about 5 kilobases to about 100 kilobases in length. The template can be circular or linear. The template can comprise all or a portion of the genomic DNA or chromosomal DNA from the source from which it is derived. The template can comprise one gene or any portion thereof. In a particular embodiment, the gene encodes for a large product, such as a polyketide synthetase or non-ribosomal peptide synthetase. The template can encode more than one gene. The template can comprise, for example DNA encoding an operon or any portion thereof, such as a desulfurization operon, a polyketide synthetase operon or a non-ribosomal peptide synthetase operon.

Methods of *in vitro* production of nucleic acid sequence are well known in the art. The template polynucleotide can be genomic sequence or fragments thereof, isolated, for example, *in vitro* using standard laboratory methods. As described herein, the methods for use with genomic DNA can also be used for other large pieces of nucleic acid, where the nucleic acid is at least about 100 kilobases in length. In particular embodiments, the double-stranded template is at least about 350 kilobases, at least about 10 megabases or at least about 20 megabases in length. For example, genomic DNA can be isolated using cell lysis in agarose plugs. Genomic fragments can be ligated into appropriate vectors, *e.g.*, Bacterial Artificial Chromosome (BAC) vectors, Yeast Artificial Chromosome (YAC) vectors, and the like, prior to application of the method of the present invention. Ligation into artificial chromosomes or other suitable vectors is especially useful for larger genomic fragments. In addition, to minimize or prevent breakage of the genomic DNA, the steps described herein can be performed in agarose plugs

The nucleic acid used in the present invention can be isolated from an organism, such as a eubacterial, archeal, eukaryotic or viral organism. These organisms can be amplified, enriched or isolated and grown in culture, or can be used directly from environmental sources. Environmental sources include but are not limited to soil samples, water samples from fresh water sources or salt water sources, polluted sites, waste treatment sites and sources including extreme condition sources such as permafrost sources, high altitude sources, high pressure sources and geothermal sources such as volcanic sources, hot springs and

hydrothermal vent sources. Sources of nucleic acid also include tissue or bodily fluid samples from an organism, such as human samples and include human genomic DNA. The nucleic acid of a tissue or bodily fluid sample can include nucleic acid of the organism, such as chromosomal, episomal or transcribed nucleic acid, or can be nucleic acid of the flora, such as fungal, bacterial, viral or parasitic organisms present in the sample. The sample can further be fresh, fossil or archival. Methods of choosing and/or isolating nucleic acids from suitable sources of nucleic acid are well known in the art.

Typically, the template is selected such that it is related to the parent polynucleotides that are used to generate the population of randomly fragmented oligonucleotides. As such, the template, more particularly, the single-stranded regions of the template, allows for the assembly of parent-derived oligonucleotides which, following, for example, trimming, gap filling and ligation, result in the formation of double-stranded chimeric polynucleotide products.

The template, parent polynucleotides or donor fragments can include one or more regions with functional characteristics or structural motifs. These regions can include nucleic acid structural motifs, protein binding domains, metal binding domains, nucleic acid binding domains, domains with enzymatic activity, or fragments of these domains. These regions can include ribozymes, deoxyribozymes, promoters, enhancers, origins of replication, open reading frames, or fragments thereof. These regions can encode aptamers, wherein aptamers are small single- or double-stranded DNA or RNA molecules that bind specific molecular targets (Bock *et al.*, *Nature 355*:564-566, 1992; Ellington and Szostak, *Nature 346*:818-822, 1990; and, Werstuck and Green, *Science 282*:296-298, 1998).

The template, parent polynucleotides or donor fragments of the present invention can also include regions of sequence that are not known to have any particular function. These regions can be selected from any known source of nucleic acid sequence, including sequences synthesized *in vitro*, or these regions can be of random or partially random sequence. Partially random sequences can be generated by synthesizing a polynucleotide based on a known sequence, except that a portion of the sequence is randomized (*e.g.*, randomizing the last 50 nucleotides), or wherein

certain positions within the sequence are randomized (*e.g.*, randomizing particular codon(s) of a coding sequence) or wherein certain bases are randomized (*e.g.*, randomizing all adenines). These regions can further encode proteins or domains of proteins including folding structures or structural motifs; binding domains such as protein binding domains, metal binding domains, co-factor binding domains, lipid binding domains and nucleic acid binding domains; domains with enzymatic function; sites for allosteric or competitive inhibition and the like; or fragments of these domains. These regions can also include amino acid sequences that are not known to have any particular function or can be randomized amino acid sequence.

The oligonucleotides are typically derived from parent polynucleotides, wherein the parent polynucleotides have been randomly fragmented. The resulting population of oligonucleotides is allowed to hybridize to the single-stranded regions of the template to ultimately form double-stranded chimeric polynucleotides. In one embodiment, the population of oligonucleotides is generated using chemical, physical or enzymatic techniques. Chemical techniques of fragmenting parent polynucleotides in order to create a fragmented population of oligonucleotides can include techniques that use pH extremes, hydroxyl radical formation, chemical radical formation, chemical catalysis or a combination thereof. Methods of fragmenting parent polynucleotides by chemical techniques can be used to generate defined or undefined ends. Techniques are well known in the art such that polynucleotides can be hydrolyzed after defined bases (*e.g.*, only after guanines), or hydrolyzed to generate undefined termini. For example, exposure of polynucleotides to extreme pH (*e.g.*, acidic pH or basic pH) can generate fragments with undefined termini. Additionally, hydroxyl radicals (*e.g.*, generated using Fenton or Udenfriend reagent) react with the deoxyribose in DNA, resulting in cleavage of the DNA strand. The result is near uniform cleavage at any base within a target polynucleotide, and the frequency of cleavage can be regulated.

In addition to generating fragmented oligonucleotides by chemical techniques, physical techniques, such as heating, freezing, using ionizing radiation and shearing, for fragmenting parent polynucleotides to produce a fragmented population of oligonucleotides can be employed.

Yet another approach to creating a fragmented population of oligonucleotides involves the use of enzymatic techniques. These methods can include the use of any suitable enzyme such as a nucleic acid polymerizing enzyme or a nuclease. For example, a polymerase can be used to synthesize oligonucleotides of variable length. Where fragments are generated by parent polynucleotide-dependent synthesis, conditions of synthesis can be chosen such that the polymerase arbitrarily falls off the polynucleotide or otherwise terminates synthesis at arbitrary points along the polynucleotide. This approach would allow for oligonucleotides to be generated with arbitrary sequence alterations (*e.g.*, "error-prone" methods). Another method for generating a fragmented population of oligonucleotides uses polymerases that are known to have exonuclease activity under conditions permitting exonuclease activity. Such enzymes include, for example, T4 DNA polymerase, PolI and PolIII.

Another method for enzymatically generating a population of oligonucleotides with undefined termini involves removing bases or generating adducts in an oligonucleotide using techniques well known in the art. For example, specific bases in oligonucleotides can be removed or adducted by many well known chemical methods to result in either abasic sites or chemically altered bases. These sites can be produced, for example, between 15 and 5000 bases apart (Kunkel *et al.*, *Meth. Enzymol. 154*:367-382, 1987). Strand cleavage of the phosphodiester bond at those modified sites can then be effected using chemicals such as piperidine, or enzymes such as abasic lyases.

Still another enzymatic method for creating a fragmented population of oligonucleotides with defined ends use endonucleases having sequence-specific recognition sites. Such enzymes are known as "restriction endonucleases" and are commercially available. A fragmented population of oligonucleotides can be generated by performing a limited or incomplete digestion of the parent polynucleotides. Additionally, oligonucleotides having undefined termini can be generated by using non-specific endonucleases such as mung bean exonuclease, S1 nuclease or DNase I. In another embodiment of generating oligonucleotides having

undefined ends, exonucleases such as Exo III or Exo VII can be used to non-specifically trim oligonucleotide sequences.

The fragmented population of oligonucleotides can include oligonucleotides wherein oligonucleotides in the population contain random or partially random sequence. The population of fragments can include molecules generated using any one of the above described methods or combinations thereof.

The term "random" as used herein is intended to reflect an absence of preselection. Such absence can be of any degree; it need not be a total absence of preselection, nor does the term indicate a requirement for an absence of preference or bias. The term can be used to describe populations of oligonucleotides, sequences, events, processes, states or conditions, or other such terms. Such compositions can range over a span of values and any one component can occupy any of these values. For example, a population of oligonucleotides which is generated by the digestion of two genes with a restriction enzyme is a "random population" when the particular oligonucleotides formed by the process are not preselected, for example, during a partial digestion. This is true even when the gene sequences are known and the restriction enzyme preferentially cleaves a particular site. Sequences can be random if at least one position in the sequence is not specifically defined (for example, if at least one position of an oligonucleotide could be and is either one of two or more nucleotides). The randomly fragmented population of oligonucleotides can include oligonucleotides wherein a portion of the oligonucleotides comprise random or partially random sequence as described herein.

In one embodiment of the present invention, the population of oligonucleotides includes oligonucleotides from about 5 to about 50,000 nucleotides in length. In a more particular embodiment, the population of oligonucleotides includes oligonucleotides from about 10 to about 10,000 nucleotides in length. In another embodiment, the population of oligonucleotides includes oligonucleotides that are about 15 to about 5,000 nucleotides in length. In another embodiment, the population of oligonucleotides includes oligonucleotides that are about 20 to about 2,500 nucleotides in length. In another embodiment, the population of oligonucleotides includes oligonucleotides that are about 25 to about 1,000

nucleotides in length. In another embodiment, the population of oligonucleotides includes oligonucleotides that are about 40 to about 200 nucleotides in length.

In one embodiment of the present invention, the population of oligonucleotides include molecules generated using any one of the above described methods or combinations thereof.

The population of oligonucleotides can include single- or double-stranded molecules. Where the oligonucleotides are double-stranded molecules, they are denatured prior to hybridization with the template. Methods of denaturing and annealing oligonucleotide sequences, are well known in the art. In one embodiment, the oligonucleotides are single-stranded. In another embodiment, the oligonucleotides of the population share minimal complementarity with each other. Oligonucleotides that share minimal complementarity with each other can be produced, for example, as described in copending United States Application No: 09/514,660 filed on February 29, 2000 or can be produced as described in "Generating Single-Stranded Oligonucleotide Librarics With Minimal Complementarity And Uses Therefor" by Joseph J. Arensdorf and Wayne M. Coco, Attorney Docket No.: 1405.2016-001, filed of even date herewith and incorporated herein by reference in its entirety.

It is clear that one of skill in the art can skew the availability of a given oligonucleotide to hybridize a template by including an oligonucleotide capable of hybridizing to said oligonucleotide. In one embodiment, a region of the template is left unhybridized by donor fragments by providing oligonucleotides complimentary to a specific template sequence. In this manner, a region of the template can be specifically retained in the resultant chimeric molecule. Conversely, defined oligonucleotides can be added in greater quantities to the population of oligonucleotides in order to preferentially hybridize the defined oligonucleotides to the template at particular regions or positions in order to introduce desired mutations or in order to protect sequences on the template from changes that might be introduced by the arbitrarily fragmented population of oligonucleotides.

The present invention allows oligonucleotides of interest to be incorporated into a larger molecule to form a double-stranded chimeric polynucleotide. In one

embodiment, polynucleotides that are not otherwise easily manipulated (*e.g.*, large polynucleotide chains), can be separately manipulated as oligonucleotides and rejoined by hybridizing the oligonucleotides to a template. For example, random mutagenesis using PCR is most effective on smaller DNA fragments, such as those 1 kilobase or less in length. A large polynucleotide can be cleaved into fragments of about one kilobase, randomly mutagenized using PCR, and then denatured. Denatured and mutagenized fragments can be hybridized to a template and ligated as described herein. The template can be derived from the original polynucleotide, or can be modified as described herein. For example, the template itself can be mutagenized or can have added or deleted regions or domains as compared to the starting polynucleotide.

It is clear to one of skill in the art that the method of the present invention can be carried out under a range of hybridization conditions using wash conditions with low to high stringencies. Conditions can be selected based on the amount of similarity or differences between the oligonucleotides and the template. In one embodiment of the present invention, the oligonucleotides are hybridized or annealed to the template or templates under conditions of low stringency.

A general description of stringency for hybridization and wash conditions is provided by Ausubel, F.M. *et al.*, *Current Protocols in Molecular Biology*, Greene Publishing Assoc. and Wiley-Interscience 1987, & Supp. 49, 2000, the teachings of which are incorporated herein by reference. Factors such as probe length, base composition, percent mismatch between the hybridizing sequences, temperature and ionic strength during hybridizations, reactions and washes influence the stability of nucleic acid hybrids. Thus, stringency conditions sufficient to allow hybridization of oligonucleotides to the template, can vary significantly and still allow for the generation of at least one chimeric polynucleotide. Typically, adjusting hybridization and wash conditions is done by, for example, adjusting the ionic strength of the reaction mixture or adjusting the temperature at which the hybridization is performed. In addition, certain purified proteins, such as the *E. coli* RecA protein, aid in homologous base pairing and can be included to facilitate hybridization of polynucleotide strands.

While not wishing to be bound by theory, typically, when two fragments anneal to form a hybridization complex, one or two single-stranded termini remain. These single-stranded termini can anneal to additional fragments from the mixture by altering hybridization conditions to favor the annealing of multiple fragments in a hybridization complex. To facilitate the hybridization of fragments having low homology, the donor and scaffold fragments can be allowed to anneal (hybridize) at 50°C. In another embodiment, the donor and scaffold fragments can be allowed to anneal at 60°C or at 70°C. To facilitate the hybridization of multiple donor fragments and scaffold fragments in a hybridization complex, the donor and scaffold fragment mixture can be held at the annealing temperature for at least about 30 seconds. In another embodiment, the donor and scaffold fragment mixture can be held at the annealing temperature for at least about 1 minute, 2 minutes, 5 minutes, 15 minutes, 30 minutes, 1 hour, 5 hours, 10 hours or 24 hours. Combinations of annealing temperature and incubation time at the annealing temperature can be used to facilitate the formation of hybridization complexes comprising multiple donor and scaffold fragments.

Alternatively, conditions for stringency are as described in WO 98/40404, the teachings of which are incorporated herein by reference. In particular, examples of "highly stringent," "stringent," "reduced," and "least stringent" hybridization and washing conditions are provided in WO 98/40404 in the Table on page 36. Examples of stringency conditions are shown in the table below which is from WO 98/40404. Highly stringent conditions are those that are at least as stringent as, for example, conditions A-F; stringent conditions are at least as stringent as, for example, conditions G-L; and reduced stringency conditions are at least as stringent as, for example, conditions M-R.

Stringency Condition	Polynucleotide Hybrid	Hybrid Length (bp)[2]	Hybridization Temperature and Buffer[1]	Wash Temperature and Buffer[1]
A	DNA:DNA	≥ 50	65°C; 1xSSC -or- 42°C; 1xSSC, 50% formamide	65°C; 0.3xSSC
B	DNA:DNA	<50	T_B*; 1xSSC	T_B*; 1xSSC
C	DNA:RNA	≥ 50	67°C; 1xSSC -or- 45°C; 1xSSC, 50% formamide	67°C; 0.3xSSC
D	DNA:RNA	<50	T_D*; 1xSSC	T_D*; 1xSSC
E	RNA:RNA	≥ 50	70°C; 1xSSC -or- 50°C; 1xSSC, 50% formamide	70°C; 0.3xSSC
F	RNA:RNA	<50	T_F*; 1xSSC	T_F*; 1xSSC
G	DNA:DNA	≥ 50	65°C; 4xSSC -or- 42°C; 4xSSC, 50% formamide	65°C; 1xSSC
H	DNA:DNA	<50	T_H*; 4xSSC	T_H*; 4xSSC
I	DNA:RNA	≥ 50	67°C; 4xSSC -or- 45°C; 4xSSC, 50% formamide	67°C; 1xSSC
J	DNA:RNA	<50	T_J*; 4xSSC	T_J*; 4xSSC
K	RNA:RNA	≥ 50	70°C; 4xSSC -or- 50°C; 4xSSC, 50% formamide	67°C; 1xSSC
L	RNA:RNA	<50	T_L*; 2xSSC	T_L*; 2xSSC
M	DNA:DNA	≥ 50	50°C; 4xSSC -or- 40°C; 6xSSC, 50% formamide	50°C; 2xSSC
N	DNA:DNA	<50	T_N*; 6xSSC	T_N*; 6xSSC
O	DNA:RNA	≥ 50	55°C; 4xSSC -or- 42°C; 6xSSC, 50% formamide	55°C; 2xSSC
P	DNA:RNA	<50	T_P*; 6xSSC	T_P*; 6xSSC
Q	RNA:RNA	≥ 50	60°C; 4xSSC -or- 45°C; 6xSSC, 50% formamide	60°C; 2xSSC
R	RNA:RNA	≤50	T_R*; 4xSSC	T_R*; 4xSSC

[2]: The hybrid length is that anticipated for the hybridized region(s) of the hybridizing polynucleotides. When hybridizing a polynucleotide to a target polynucleotide of unknown sequence, the hybrid length is assumed to be that of the hybridizing polynucleotide. When polynucleotides of known sequence are hybridized, the hybrid length can be determined by aligning the sequences of the polynucleotides and identifying the region or regions of optimal sequence complementarity.

[†]: SSPE (1xSSPE is 0.15M NaCl, 10mM NaH_2PO_4, and 1.25mM EDTA, pH 7.4) can be substituted for SSC (1xSSC is 0.15M NaCl and 15mM sodium citrate) in the hybridization and wash buffers; washes are performed for 15 minutes after hybridization is complete.

*T_B - T_R: The hybridization temperature for hybrids anticipated to be less than 50 base pairs in length should be 5-10°C less than the melting temperature (T_m) of the hybrid, where T_m is determined according to the following equations. For hybrids less than 18 base pairs in length, T_m(°C) = 2(# of A + T bases) + 4(# of G + C bases). For hybrids between 18 and 49 base pairs in length, T_m(°C) = 81.5 + 16.6($\log_{10}$[Na^+]) + 0.41(%G+C) - (600/N), where N is the number of bases in the hybrid, and [Na^+] is the concentration of sodium ions in the hybridization buffer ([Na^+] for 1xSSC = 0.165 M).

It is clear to one of ordinary skill in the art that the contacting and hybridization steps can be optimized using any suitable method of optimization that is established in the art of hybridization. These include, but are not limited to, techniques that increase the efficiency of annealing or hybridization from complex mixtures of polynucleotides (*e.g.*, PERT; *Nucleic Acids Research 23*:2339-2340, 1995) or hybridization in different formats (*e.g.*, using an immobilized template or using microtiter plates; *Analytical Biochemistry 227*:201-209, 1995).

Any parent polynucleotide with sufficient sequence similarity to the template can be used to generate the oligonucleotides of the present invention. As defined herein, "sufficient sequence similarity" means that the sequence of the oligonucleotide need not reflect the exact sequence of the template. Conditions are chosen to allow such sequences (and those having low similarity or similar sequences interrupted with dissimilar sequences) to hybridize the template, such that chimeric polynucleotides are formed. For example, non-complementary bases or insertions or deletions can be interspersed in sequences.

Where flaps (unhybridized termini), gaps (single-stranded regions) and/or nicks occur in the hybridized complex, they are trimmed, filled and ligated, respectively as described in copending Serial No: 09/514,660. In a particular embodiment, immediately adjacent oligonucleotides are ligated to each other. The term "adjacently hybridized" is used herein to describe the relative positions of two donor fragments hybridized to the same single-stranded region, at positions such that only single-stranded sequence is contained between the two fragments. The term "immediately adjacently hybridized" is used herein to describe adjacently

hybridized donor fragments that abut each other, *e.g.*, no intervening single-stranded sequence is contained between the two hybridized fragments.

Typically, a trimming, polymerization, ligation (TPL) step follows the contacting and hybridization of the population of oligonucleotides to the template. The TPL step typically includes trimming flaps, polymerization to fill in gaps between adjacently hybridized oligonucleotides, and ligation to join immediately adjacently hybridized oligonucleotides and to join hybridized oligonucleotides to immediately adjacent template nucleic acid.

The utility of trimming flaps is realized because, in certain cases, the population of oligonucleotides can hybridize to the template or templates such that at least one terminus of at least one of the hybridized oligonucleotides is unhybridized. The term "flaps" is used herein to describe the unhybridized terminal portion of an oligonucleotide otherwise hybridized to a template. "Loops" is used to describe unhybridized internal portions. The "trimming" of flaps, used herein to refer to a process of removing the unhybridized terminal portions, leaving the hybridized portion of the oligonucleotide and intervening single-stranded regions intact, can be incorporated into the method of the present invention. Flaps can be trimmed enzymatically, *e.g.*, utilizing polymerases with single-stranded exonuclease activity or other single-stranded endonucleases or exonucleases, or chemically. The step of trimming flaps can be performed prior to or concurrently with the additional steps of polymerization and ligation. In some embodiments, loops can also be trimmed.

The polymerization step is utilized to fill in "gaps" between adjacently hybridized oligonucleotides and between hybridized oligonucleotides and adjacent template nucleic acid. Depending on specific hybridization capabilities, oligonucleotides can hybridize to the template or templates such that regions of the template remain unhybridized, *i.e.*, "gaps" are created. Such gaps could prevent the final formation of a template-length chimeric polynucleotide, so a polymerization step is used to fill in the gaps.

Polymerization can be achieved either chemically or enzymatically. For example, gaps between adjacently hybridized oligonucleotides can be filled using a

suitable nucleic acid polymerizing enzyme, *e.g.*, a "polymerase". Suitable polymerases are commercially available. In one embodiment, gaps are filled in using prokaryotic, eukaryotic or viral polymerases. The polymerase can be thermostable or not thermostable. The polymerases can optionally have proof reading ability. Suitable polymerizing enzymes include T4 DNA polymerase, *Taq* DNA polymerase, *Pfu* DNA polymerase, Pol I, Klenow and Klenow 3'→5'^{exominus} (New England BioLabs, Beverly, MA).

Control of enzymatic polymerization can be achieved, for example, by affecting the polymerase, *e.g.*, using a polymerase with altered processivity, or by affecting the template which is used by the polymerase during polymerization.

In the method of the present invention, gaps between adjacently hybridized oligonucleotides are separated by about 1,000 to about 100,000 template nucleotides. In other embodiments, the adjacently hybridized oligonucleotides are separated by about 500 to about 10,000 template nucleotides; less than 1,000 template nucleotides; less than 250 template nucleotides; less than 50 template nucleotides; or are separated by less than 25 template nucleotides.

In another embodiment, gaps are filled in *in vivo*, wherein complexes containing oligonucleotides hybridized to a template are inserted or transformed into a suitable host cell. For the purposes of the present invention, the gaps can be filled with or without the introduction of "errors" in comparison to the template.

In the method of the present invention, hybridized oligonucleotides are ligated. The hybridized oligonucleotides to be ligated are hybridized immediately adjacent to each other or immediately adjacent to template nucleic acid. The ligation can be performed using a suitable ligase or using one or more ligases. Suitable ligases include thermostable and non-thermostable ligases and include, but are not limited to, T4 DNA ligase, DNA ligase I, *Taq* ligase and *Tth* ligase. In another embodiment, ligation is performed using chemical means.

The method of the present invention typically generates chimera of at least about 50% of the length of the target sequence after only one round or iteration of the method. In more particular embodiments, the method generates chimera of at least about 60, 70, 80, 90 and 100% of the length of the target sequence after only

one round of the method. The method of the present invention can further include repeating the method using at least one chimeric polynucleotide or fragment thereof as the hybridization template or source of hybridizing oligonucleotides.

The final chimeric product is a heteroduplex that typically does not contain a contiguous, full-length template. It is therefore unnecessary to modify the template strand to facilitate its removal. The heteroduplex product can be amplified using standard amplification techniques to generate homoduplex chimera or can be cloned and introduced into an organism using standard cloning and transformation techniques upon which replication *in vivo* will generate homoduplex chimeric molecule. The two strands of each chimeric heteroduplex can be segregated during replication in the transformed host cell and each strand of the heteroduplex carried forward in separate progeny. Where chimeric regions have been generated on both strands of the template, a "chimeric heteroduplex" is created. The two strands of the chimeric heteroduplex can each contain different chimeric sequences. The two chimeric strands of the heteroduplex can be segregated during subsequent replication of transformants. Selection can be used to enrich various improved chimeric progeny from unimproved progeny or parental carry over. Auxotrophic or antibiotic sensitive host strains can be used to minimize the occurrence of non-transformants in the *in vivo* library and to minimize the propagation of segregants containing only the host genome.

The resulting library can be transformed into cells of interest using standard methods which are apparent to one skilled in the art, *e.g.*, protoplast-liposome fusion technology, electroporation, or yeast spheroplast transformation. In the case of plants, chimeric polynucleotides as large as 150 kb can be introduced into the plant genome using bacterial artificial chromosome (BAC) vectors and transformation *e.g.*, via *Agrobacterium*.

In still another embodiment of the present invention, the chimeric polynucleotides are selected or screened based on alterations of specific properties, *e.g.*, nucleotide structure, nucleotide function, altered enzymatic activities of proteins encoded by the chimeric polynucleotide, altered structural functions of proteins encoded by the chimeric polynucleotide, altered regulatory functions of

proteins encoded by the chimeric polynucleotide, etc., or a combination thereof. Subsequent selection and amplification of the chimeric polynucleotide allows for the *in vitro* or *in vivo* directed evolution of biological molecules such as nucleic acid or polypeptides. This method for directed evolution would aid in the improvement of such molecules for use, for example, in medical therapies, as reagents in molecular biology, and in industry.

The present invention is particularly useful for evolving industrially or medically useful molecules biochemical pathways, or regulatory sequence. The resulting chimeric polynucleotides can comprise a regulatory sequence (*e.g.*, promoter, aptamer, catalyst, enhancer or other regulatory element) or can encode a useful gene product, or can encode enzymes in a useful biosynthetic pathway. The chimeric polynucleotides can be or encode molecules that are more active under desired conditions to have altered or enhanced specificity, mutagenicity or fidelity. For example, desired conditions include conditions to which the reference molecule, oligonucleotide, template, or polypeptide encoded therein is not typically exposed or otherwise extreme conditions. Extreme conditions could include high or low temperature, extreme high or low pH, extreme ionic strength, extreme solvent conditions such as organic solvent conditions, or a combination of two or more of these conditions.

One of skill in the art can readily select or design a template to encode the molecule of interest to be evolved according to the method of the present invention. Even where biospecific genes have not been specifically identified or sequenced, the genome of the microbe or organism known to produce the gene product can be used, for example, as a template with donor fragments from genes or genomes containing biosynthetic genes for the production of similar products. The template can also include introns.

Examples of industrially useful molecules include enzymes that synthesize polyketides, catalyze non-ribosomal peptide synthesis, transform small molecules, hydrolyze substrates, replace steps in organic synthesis reactions or degrade pollutants such as aromatic hydrocarbons (*e.g.*, benzene, xylene, toluene and naphthalene), polychlorinated biphenyls and residual herbicides and pesticides.

Catabolic pathways can be evolved using the present invention such that enzyme pathways are produced that degrade manmade pollutants that otherwise are not or only slowly catabolized. Polynucleotides encoding such enzymes or fragments of coding regions can be used in the present invention as either the template, parent polynucleotides, a reference molecule to which chimeric polynucleotide products are compared, or combinations thereof.

Examples of industrially or medically useful polypeptides or polynucleotides are well known in the art. Medically useful molecules include "bioactive" molecules, used herein to include peptides; proteins; polysaccharides and other sugars; lipids; and nucleic acid sequences, such as genes, and antisense molecules. Nucleic acid encoding enzymes that produce, modify or degrade polysaccharides, other sugars or lipids can be used as the template, oligonucleotides or reference polynucleotide. Specific examples of bioactive molecules include, but are not limited to, insulin, erythropoietin, interferons, colony stimulating factors such as granulocyte colony stimulating factor, growth hormones such as human growth hormone, LHRH analogs, LHRH antagonists, tissue plasminogen activator, somatostatin analog, Factor VIII, Factor IX, calcitonin, dornase alpha, polysaccharides, AG337, bone inducing protein, bone morphogenic protein, brain derived growth factor, gastrin 17 immunogen, interleukins such as IL-2, PEF superoxide, permeability increasing protein-21, platelet derived growth factor, stem cell factor, thyrotropin and somatomedin C.

Methods for measuring activity of hormones, interleukins, growth factors and angiogenesis inhibitors and the like under desired conditions are well known in the art. One of ordinary skill in the art can readily determine the activity of the hormone, interleukin, growth factor, angiogenesis inhibitor or antibiotic encoded by the chimeric polynucleotide produced by the present invention and select those having the desired characteristics. Examples of medically useful molecules to be evolved according to the present invention also include enzymes that synthesize drugs, antibiotics, vitamins or co-factors. Other examples include vectors and genes for gene therapy. In addition, molecules that have desired therapeutic effect can be altered to lessen toxicity, antigenicity or other side effects.

Methods for determining activity under desired conditions include standard methods well known in the art. One of ordinary skill in the art can readily determine the activity of an enzyme encoded by a chimeric polynucleotide and select those polynucleotides that encode enzymes that have desired characteristics. Enzymes include but are not limited to fermenting enzymes, proteases, lipases, oxidoreductases such as alcohol dehydrogenase, polymerases, hydrolases and luciferase.

In one embodiment of the present invention, a chimeric polynucleotide is generated wherein one or more characteristics of the product molecule is different with respect to at least one reference polynucleotide. The difference in the chimeric polynucleotide can include a nucleotide change and/or amino acid changes in the encoded polypeptide in comparison to the reference polynucleotide, polypeptide or fragment thereof. The reference polynucleotide, polypeptide or fragment thereof can be the template or fragment, or can be a molecule related to the template used for comparison. For example, where the template is a non-functional version of a polynucleotide of interest or polypeptide encoded therein, then a reference molecule can be used for comparison to chimeric polynucleotides generated. The reference molecule can be a family member of the gene or gene product of interest, such as a homologous gene, or fragment thereof. One of skill in the art can readily choose a reference molecule based on the templates and oligonucleotides of interest used to generate the chimeric polynucleotides.

The characteristics to be altered according to the present invention include, but are not limited to, structural motif, stability, half-life, enzymatic activity, enzyme specificity, binding affinity, binding specificity, toxicity, antigenicity, interaction with an organism or interaction with components of an organism of the polynucleotide or the encoded polypeptide. A functional characteristic can be altered according to the present invention such that the activity of said functional characteristic is enhanced at a higher or lower temperature compared to a reference molecule. Furthermore, said functional activities can be enhanced in various physical or chemical environments as described above. Methods for measuring, selecting and screening these characteristics are well known in the art.

Structural motifs for proteins include, for example, α-helices, beta-sheets, solvent exposed loops, leucine zippers, β-barrel scaffolds and the like. Structural motifs for polynucleotides include, for example, quadraplexes, aDNA, bDNA, zDNA, triple helices, stem loops, hairpins, protein binding sites and the like. Examples of regions are provided above. Methods for determining these motifs are well known in the art. In one embodiment, alteration of the characteristic includes an enhancement of the characteristic. In another embodiment, alteration of the characteristic includes a reduction in the characteristic.

In one embodiment of the present invention, the chimera is cloned prior to selection or screening. Methods of cloning polynucleotides are well known in the art. Alternatively, the chimera can be selected or screened *in vitro* or *in vivo* prior to cloning.

The present invention allows the generation of at least one chimeric polynucleotide based in part on hybridization with at least one template. The chimeric polynucleotides are different from any single template used to generate the chimeric polynucleotide. Based on the method of the present invention, the differences can include, for example, an additional region, wherein the region is not present in the template. The additional region can be derived from an existing source of polynucleotides, or a modified form thereof or can be a partially or completely random sequence. The additional region or regions can be present at either terminus of the resultant chimeric polynucleotide or can be present within the chimeric polynucleotide. Thus, the chimeric polynucleotide of the present invention can be longer than the template. In another embodiment, the chimeric polynucleotide can include an altered version of a region that is present in the template. The region can be the same length as the region in the hybridization template or can be longer or shorter than the region in the hybridization template. Thus, the chimeric polynucleotide can be the same size, longer or shorter than the template.

The invention will be further illustrated by the following non-limiting example.

EXEMPLIFICATION

This invention can be applied to DNA target sequences over a wide range of sizes. Chimeragenesis can be performed on whole genomes, on large fragments thereof, or on target sequences that have been cloned into vectors, *e.g.*, BACs, YACs, plasmids, etc. The invention can be applied to the whole genome of bacteria, yeast, fungi, or other organisms of interest. The whole genome can be isolated intact, chimerized intact, and transformed back into an appropriate host. Alternatively, the whole genome can be fragmented or fragmented and cloned into vectors both before or after the chimeragenesis step. The invention can also be applied to previously cloned DNA targets of any size that reside on vectors.

Isolation of target sequences and transformation procedures are numerous and well known to those skilled in the art. Methods exist for the manipulation of intact polynucleotides up to several megabases in length, but it is advisable to work in agarose plugs if the target sequence length is over 100 kb to prevent DNA breakage. These examples describe chimeragenesis methods both for targets entrained in agarose plugs and for targets in solution.

Example 1:

Isolation of DNA and chimeragenesis in agarose plugs

For chimeragenesis of whole genomic DNA, the following procedure has been developed. Intact whole genomic DNA is first isolated from the organism of interest. Standard methods for isolation of genomic DNA from mammalian, plant, yeast, and bacterial cells are known to those skilled in the art. Isolation of genomic DNA by cell lysis in agarose plugs generates intact DNA that can be manipulated with minimal shearing. As an example, agarose, *e.g.*, InCert Agarose (BioWhittaker Molecular Applications, Inc., Rockland, ME.) is prepared in TE to 1-2% and held at 50°C. Cells are grown to mid-log phase, washed in a Tris-NaCl buffer, and resuspended in 1 ml Tris-NaCl buffer per 10 ml original growth culture. The cells are warmed to 37°C and mixed 1:1 with the InCert agarose. The cell/agarose suspension is quickly poured into a mold and placed on ice to solidify for approximately 10 minutes. The cell/agarose suspension can be poured into a

sterile 5 ml syringe with the end plugged that serve as a mold. The end of the syringe is removed and the solidified agarose is gently pushed out of the syringe and cut into thin slices with a sterile razor. The volume of the slices prepared will vary depending on the goals of the experiment, but slices approximating 100 µl volume are typically used. Many slices are produced and are manipulated as replicates for subsequent procedural steps.

Many gel slices (plugs) can be combined in a small beaker for the lysis steps. Lysis is typically carried out as follows. Two volumes (per total volume gel plugs) of a lysozyme lysis buffer (lysozyme 1 mg/ml; RNase 20 µg/ml) are added to the plugs and incubated with shaking at 37°C overnight. The lysozyme lysis buffer is aspirated off and replaced with 1 volume of Proteinase K solution (0.5 M EDTA, 1% N-lauroylsarcosine, and 1 mg/ml Proteinase K). The plugs are incubated with shaking at 50°C for 24 hours, the Proteinase K solution is replaced with fresh Proteinase K solution and the plugs are incubated another 24 hours at 50°C with shaking. The solution is aspirated and the plugs are incubated in five volumes of TE buffer with 1mM phenylmethylsulfonyl fluoride at 37°C for 2 hours. The solution is changed with a fresh five volumes and incubated for two hours to overnight. The solution is then aspirated, five volumes TE buffer are added, and the plugs are incubated for 2 hours at 37°C. A second wash with five volumes of TE is performed. The plugs contain genomic DNA at this stage and can be stored for weeks or months.

For the chimeragenesis steps, individual plugs, typically 100 µl volume each, are placed in microcentrifuge tubes. Because the exact conditions for chimeragenesis will change based on enzyme lot, DNA source, agarose lot, etc., the chimeragenesis methods are typically performed on many replicate plugs simultaneously using different conditions in parallel. This allows routine optimization of conditions. In some cases, where these parameters do not vary widely, a single reaction condition may be sufficient.

A set of plugs is first treated with DNase I. The density of nicks generated by DNase I can be controlled by adjusting the digestion time, pH, temperature, divalent cation, and enzyme concentration or additives that alter the kinetics of

DNase I. The density of nicks can be varied using routine optimization of digestion time, pH, temperature, divalent cation and enzyme concentration or additives that alter the kinetics of DNase I. To each plug, two volumes of DNase buffer (100 mM sodium acetate pH 5, 5 mM $MgCl_2$, 100 µg/ml BSA) are added. It is important that no manganese be in the buffer as manganese promotes double stranded nicks. Each tube is incubated for 15 minutes. The buffer is aspirated, and two volumes of DNase buffer containing 0.1 µg/ml DNase I is added. The samples are incubated with gentle shaking at 15°C. Several replicates of each are stopped at intervals between 5 and 30 minutes. The digests are stopped by quickly aspirating the DNase buffer, adding two volumes of the Proteinase K solution, and incubating with gentle shaking at 50°C. After 20 minutes, the solution is aspirated and the plugs are incubated in five volumes of TE buffer with 1mM phenylmethylsulfonyl fluoride at 37°C for 20 minutes. The solution is changed for a fresh five volumes of exonuclease buffer (50 mM Tris, pH 8.5 mM $MgCl_2$ 5 mM DTT, 50 µg/ul BSA) and incubated with gentle shaking for 30 minutes. The buffer is then aspirated, five volumes of fresh exonuclease buffer is added, and the plugs are incubated for another 30 minutes.

At this point, one replicate plug from each time point can be used to determine the density of nicking as follows. The test plugs are treated by standard methods to liberate the DNA from the agarose, *e.g.*, agarase digestion, heating, or diluting in NaI. The DNA is resuspended in a denaturing buffer, *e.g.* urea, incubated at 95°C for several minutes, quickly loaded onto a denaturing agarose gel, and electrophoresed. The size of the single stranded bands/smears visualized on the agarose gel will allow an estimation of the nicking density. The replicate set(s) with the desired nicking density are processed further. If none of the samples display the desired nicking density, the DNase digestion is repeated on fresh plugs using modified digestion conditions. DNase digestion should be optimized for any given batch of DNase and for each DNA template. To optimize the reaction, the nicking density can be, for example, decreased by decreasing enzyme concentration, decreasing the reaction temperature, decreasing the pH, or decreasing the reaction

time, or adjusting the amounts of divalent cation or other additives that alter the kinetics of DNase I.

DNA in plugs is then digested with exonuclease III to produce single stranded gaps. The gap length is carefully controlled so that gaps rarely cross nicks on adjacent strands and break the duplex. The processivity of the exonuclease digestion can be controlled by adjusting the digestion time, pH, temperature, sodium chloride concentration, or enzyme concentration. For each plug to be digested, the buffer is aspirated, and two volumes of fresh exonuclease buffer containing 100 units/ml exonuclease III are added. The samples are incubated with gentle shaking at 37°C. Samples of each condition are stopped at several intervals between 2 and 30 minutes. The digests are stopped by quickly aspirating the buffer, adding two volumes of Proteinase K solution, and incubating with gentle shaking at 50°C. After 20 minutes, the solution is aspirated and the plugs are incubated in five volumes of TE buffer with 1mM phenylmethylsulfonyl fluoride at 37°C for 20 minutes. The solution is changed for a fresh five volumes of annealing buffer (20 mM Tris, pH 8, 25mM potassium acetate, 10mM DTT, 0.5 to 10 mM magnesium acetate, 1 mM NAD, 0.1% Triton X-100™) and incubated with gentle shaking for 30 minutes. The buffer is then aspirated, five volumes of fresh annealing buffer is added, and the plugs are incubated for another 30 minutes.

At this point one replicate plug from each time point can be used to determine the processivity of the exonuclease digestion. The test plugs are treated by standard methods to liberate the DNA from the agarose, *e.g.*, agarase digestion, heating, or diluting in NaI. The DNA is resuspended in a denaturing buffer, *e.g.*, urea, incubated at 95°C for several minutes, quickly loaded onto a denaturing agarose gel, and electrophoresed. The size of the single stranded bands/smears visualized on the agarose gel, compared with controls that were treated with DNase I but not exonuclease, will allow an estimation of the average gap size. The replicate set(s) with the desired gap size(s) are processed further. If none of the samples displays the desired gap size the exonuclease digestion is repeated on additional plugs using varied digestion conditions. To optimize the reaction, gap size can be decreased by addition of sodium chloride up to 50 mM, decreasing

enzyme concentration, decreasing the reaction temperature, decreasing the pH, or decreasing the reaction time.

The next step is to anneal single-stranded donor fragments to the single stranded gaps in the genomic DNA. Donor fragments may be generated by the methods described herein. If the source of donor fragments is complex DNA, *e.g.*, whole genomic DNA, human chromosomal DNA, environmental DNA, the method taught in "Generating Single-Stranded Oligonucleotide Libraries With Minimal Complementarity And Uses Therefor" by Joseph J. Arensdorf and Wayne M. Coco, Attorney Docket No.: 1405.2016-001, filed of even date herewith, may be used to generate the donor fragments. Alternatively, synthetic oligonucleotides designed for sequence specific targets can be used as donor fragments or added in addition to donor fragment pools. The amount of donor fragments used depends on the specific experimental goals, but several reaction series can be conducted in parallel over a range of donor fragment concentrations that span several orders of magnitude, *e.g.*, 10 ng to 50 µg per reaction. Each plug to be annealed is preheated to 50°C, the annealing buffer is aspirated, and two volumes of fresh annealing buffer containing donor fragments (also equilibrated to 50°C) are added. The plugs are incubated at 50°C with gentle shaking for one to six hours. Depending on the goals of the experiment and the desired stringency for annealing, the concentration of magnesium acetate used in the annealing buffer can be varied from 0.5 to 10 mM. Lower magnesium concentrations help establish high stringency annealing conditions and higher magnesium concentrations help establish low stringency annealing conditions. If the appropriate stringency is not known, replicates can be processed in parallel with each of several magnesium concentrations. Alternatively, the stringency can be started high and lowered over the course of the annealing step by periodic addition of magnesium acetate to the tube during incubation. Incubation temperatures lower than 50°C can also be employed to lower annealing stringency. At the end of the annealing incubation, the buffer is aspirated and replaced with two volumes of 1X *Taq* ligase buffer.

The plugs are soaked for two hours during which time the *Taq* ligase buffer is aspirated and replaced every 30 minutes. After the final buffer replacement, 200

units of *Taq* ligase and 5 units of *Taq* polymerase are added to each tube. *Taq* polymerase is added to trim 5' flaps and the ligase is included to ligate nicks. The plugs are incubated at 50° C for 5 to 30 minutes, and then the following additions are made to each sample: 15 units of *Pfu* polymerase, 5 units of additional *Taq* polymerase, and dNTPs to 200 μM. *Pfu* polymerase is added to trim 3' -flaps and to assist in filling of large gaps. *Taq* polymerase concentration is supplemented and dNTPs are added for the purpose of filling in gaps. The *Taq* ligase already present will ligate nicks after polymerization across gaps. Alternative polymerases and ligases can be used if they are active at 50°C and do not possess features, *e.g.*, unfavorable exonuclease activity, that might interfere with the chimeragenesis process. The plugs are incubated at 50°C with gentle shaking for 1 to 4 hours. Sample plugs can be used at intervals to run on denaturing agarose gels to evaluate the extent to which trimming, polymerization, and ligation has restored complete double stranded heteroduplexes. If there is evidence that single-stranded gaps still remain, the TPL steps can be repeated with appropriate adjustments in enzyme concentration, pH, buffer composition, incubation time, or temperature.

At this stage the chimeric heteroduplexes can be transformed into appropriate host cells as described below, restricted and cloned into vectors, or used as the template for PCR of specific target genes/gene clusters.

Using standard methods known to those skilled in the art, the genomic DNA can be restricted in the plug or out of the plug and ligated into appropriate vectors and transformed into the desired host. Alternatively, the whole genomic chimera can be mobilized into a host.

This method describes the digestion of the agarose matrix with an agarase enzyme *e.g.*, Gelase (Epicentre, Madison, WI), the formation of liposomes, which contain the genomic DNA, and the fusion of the liposomes with protoplasts of the desired host organism to introduce the chimeric genome into the host. The buffer in which the plugs were incubated is aspirated and exchanged for ten volumes of the agarase enzyme buffer by shaking gently for one hour, exchanging fresh agarase buffer, and incubating another one hour. The buffer is removed and the plugs are heated in the tube to 70°C to melt the agarose. The tubes are transferred to 45°C

and after equilibration, the manufacturer's recommended amount of agarose is added. Samples are incubated for one hour at 45°C. At this stage, the sample may be pipetted with a wide bore tip, *e.g.*, a yellow tip with the tip cut off, to a MICROCON™ concentrator (Millipore, Bedford, MA) to isolate the DNA with minimal shearing for subsequent steps to generate DNA:liposome complexes. Alternatively, DNA:liposome complexes can be formed directly in the agarose sample. To generate DNA:liposome complexes, 2-25 µl of DOTAP (available from several manufacturers, *e.g.*, Boehringer-Mannheim, Indianapolis, IN.) is added to the samples and the tubes are incubated at room temperature for 15 minutes. The DNA:liposome complexes that are formed will readily fuse with protoplasts (*e.g.*, gram positive bacteria), spheroplasts (*e.g.*, gram negative bacteria and yeast), or eukaryotic cells without cell walls to release the DNA into the cytoplasm.

Protoplasts/spheroplasts of the appropriate host can be made by standard methods well known to those skilled in the art. Formation of protoplasts/spheroplasts typically involve incubating the cells with lytic enzymes (lysozyme) in the presence of osmotic stabilizers (sucrose). The exact method of protoplast formation depends on the genera of host cell to be transformed. After protoplasts are generated, they are mixed with the DNA:liposome complexes. Typically, 1 ml of protoplasts resuspended in 0.85 M sucrose is mixed with 100 µl of DNA:liposome complex (the product of one plug) and incubated at room temperature for four hours. Some methods require the addition of PEG to promote fusions. The mixture is then transferred to a rich broth, incubated at the preferred temperature of the host for several hours, and then plated onto the appropriate media for selection. Protoplasts can be UV-irradiated prior to fusion in order to disable the hosts genome and increase selection for segregants transformed with chimeric genomes. It can take several generations for the chimeric genome to be segregated. In some genera, the transformed chimeric genome can recombine with the host genome. For some applications, it may be advisable to plate transformants onto nonselective media and then to screen (or replate onto selective media) those colonies which arise.

Example II:

Chimeragenesis in solution of genomic fragments or polynucleotide targets in vectors

Chimeragenesis of polynucleotide targets less than 100 kb can be performed in solution instead of in agarose plugs. This embodiment has application to genomic fragments or polynucleotide targets residing on vectors such as BACs and plasmids. For very large vector constructs, *i.e.*,>100 kb, it is advisable to perform chimeragenesis in agarose plugs as described above to prevent DNA breakage.

The construction of and isolation of BACs, YACs, plasmids, cosmids, and other vectors is well established, and the methodologies are varied and well known to those skilled in the art.

The exact conditions for chimeragenesis can change from experiment to experiment based on enzyme lot, DNA source, agarose lot, etc. Chimeragenesis methods are typically performed on many replicate samples simultaneously using different conditions in parallel to optimize conditions using routine optimization techniques. The amount of template DNA used will depend on the goals of the experiment, but 20 fmole per replicate is typically used as a starting point. The first two steps, treatment with DNase I and treatment with Exonuclease III, are typically performed on several batches of larger amounts of DNA, and those batches with the desired nick density and gap size are subsequently split into replicates for the chimeragenesis steps.

For treatment with DNase I, each batch of template DNA is resuspended in 20-50 µl of DNase buffer (100 mM sodium acetate pH 5, 5 mM $MgCl_2$, 100 µg/ml BSA). It is important that no manganese be in the buffer as manganese promotes double stranded nicks. DNase I is then added to each batch in a small volume, *e.g.* 1 µl, to 0.01 µg enzyme per 1 µg polynucleotide, and the samples are immediately placed at 15°C to incubate. A batch is stopped at each of several intervals between 5 and 30 minutes by addition of 0.5 M EDTA and heating to 70°C for 10 minutes. At this point, a sample from each batch is used to determine the density of nicking. For each of these test samples, a denaturing agent such as urea is added, and then the test samples are incubated at 95°C for several minutes, quickly loaded onto a

denaturing agarose gel, and electrophoresed. The size of the single stranded bands/smears visualized on the agarose gel allows an estimation of the nicking density. The batch(es) with the desired nicking density is processed further, and if more template is required, the DNase step can be repeated with the best conditions. If none of the samples display the desired nicking density, the DNase digestion is repeated on fresh samples using modified digestion conditions. DNase digestion is optimized for any given batch of DNase and for each DNA template. To optimize the reaction, the nicking density can be decreased by decreasing enzyme concentration, decreasing the reaction temperature, decreasing the pH, decreasing the reaction time or adjusting the amounts of divalent cation or other additives that alter the kinetics of DNase I.

In some cases it may be desirable to introduce only one nick per vector. In that case, it may not be possible to distinguish nicked from unnicked polynucleotides on the denaturing agarose gel. An alternative method of determining nick density is to end label nick sites with a radio label, fluorescent label, or some other label. Samples can be denatured, quick-cooled, and end-labeled using methods well known in the art. The end-labeled sample can be purified and the amount of label quantified against a calibration curve. With a knowledge of the DNA quantity and vector size, the amount of labeling can be used to calculate the average number of nicks per vector.

The DNase treated template is processed in batches for digestion with Exonuclease III in order to produce single stranded gaps. The DNA in selected samples is then cleaned by any of several methods, *e.g.*, QIAGEN™ kit (Qiagen, Valencia, CA), MICROCON™ column (Millipore, Bedford, MA), etc., and the DNA is resuspended in exonuclease buffer (50mM Tris, pH 8, 5 mM $MgCl_2$, 5 mM DTT, 50 μg /μl BSA). Exonuclease III is added at 50 units enzyme per 1 μg polynucleotide to each batch and batches are incubated at 37°C. Several replicates are stopped at each of several intervals between 2 and 30 minutes. The digests are stopped by addition of 0.5 M EDTA and heating to 70°C for 20 minutes. At this point, a test sample from each batch can be taken to determine the processivity of the exonuclease digestion. For each of these test samples a denaturing agent such as

urea is added, and then the test samples are incubated at 95°C for several minutes, quickly loaded onto a denaturing agarose gel, and electrophoresed. The size of the single-stranded bands/smears visualized on the agarose gel, compared with controls that were treated with DNase but not exonuclease, will allow an estimation of the average gap size. The batch(es) with the desired gap size(s) is processed further, and if more template is required, the DNase step can be repeated with the best conditions. If none of the samples displays the desired gap size, the exonuclease digestion is repeated on fresh samples using varied digestion conditions. To optimize the reaction, gap size can be decreased by addition of sodium chloride up to 50 mM, decreasing enzyme concentration, decreasing the reaction temperature, decreasing the pH, or decreasing the reaction time.

In some cases it may be desirable to introduce only one long gap per vector. This would entail the introduction of one nick per vector as described above, followed by a long exonuclease digest. In some such cases, *e.g.*, if vector is large, it may not be possible to distinguish gapped and ungapped polynucleotides on the denaturing agarose gel. An alternative method of determining gap length is to treat a representative sample with *Taq* polymerase in the presence of labeled dNTPs. The nucleotides may be radio labeled, fluorescently labeled, or contain some other label, *e.g.*, biotin. After the gap is filled by the polymerase, the sample DNA can be purified and the amount of label quantified against a calibration curve. With a knowledge of the starting vector quantity and vector size, the amount of labeling can be used to calculate the average gap size per vector.

The next step is to anneal single stranded donor fragments to the single stranded gaps in the genomic DNA. Donor fragments can be generated by the methods described in copending United States Application No: 09/514,660. If the source of donor fragments is complex DNA, *e.g.*, whole genomic DNA, human chromosomal DNA, environmental DNA, etc., the "Generating Single-Stranded Oligonucleotide Libraries With Minimal Complementarity And Uses Therefor" by Joseph J. Arensdorf and Wayne M. Coco, Attorney Docket No. 1405.2016-001 filed of even date herewith, may be used to generate the donor fragments. Alternatively, synthetic oligonucleotide probes designed for sequence specific targets may be used

as donor fragments or added to donor fragment pools. The amount of donor fragments used will depend on the specific experimental goals, but it is recommended that several reaction series be conducted in parallel over a range of donor fragment concentrations that span several orders of magnitude. For chimeragenesis of a sample containing 1 μg of template DNA, donor fragments will typically be tested between 10 ng and 50 μg total per reaction. If an adequate template:donor ratio cannot be achieved, then the amount of template can be decreased. The size range depends on the experimental goals, but donors between 50 bp and 500 bp are typically used. It is advisable to split donor fragments into size ranges, *e.g.* 100-200 bp, and run the different size ranges as replicates in parallel.

The template DNA in selected samples is then cleaned by any of several methods, *e.g.*, QIAGEN™ kit, MICROCON™ column, etc., and then resuspended in 1X *Taq* Ligase buffer, typically to 20 μl. The template samples are supplemented with tRNA to 1 ng/μl final concentration and then heated to approximately 80°C. The donor fragments are resuspended in 1X *Taq* Ligase Buffer, heated to 94°C, and then added to the template samples in small volumes, *e.g.* 1-2 μl, to the prescribed concentrations, and then immediately placed at the annealing temperature. The annealing temperature is typically 60°C, but this temperature can be adjusted depending on the goals of the experiment and the desired stringency for annealing for different templates, *e.g.*, lower temperature for AT-rich templates. Alternatively, the magnesium acetate concentration in the *Taq* Ligase Buffer can be varied from 0.5 to 10 mM in order to fine tune annealing stringency. The sample is held at the annealing temperature for 30 minutes. If the appropriate stringency is not known, replicates can be processed in parallel with each of several different temperatures and magnesium concentrations. Alternatively, the stringency can be started high and lowered over the course of the annealing step by periodic addition of magnesium acetate and/or gradual lowering of the temperature during incubation. At the end of the annealing incubation, 40 units of *Taq* ligase and 2 units of *Taq* polymerase are added in 1X *Taq* ligase buffer. The sample is placed at 60°C and incubated for 5 minutes. The temperature of incubation is then lowered to 45°C,

and 3 units of Pfu polymerase is added with dNTPs to 100μM. The temperature is then slowly ramped from 45°C to 60°C over 30 minutes. The purposes of each of the enzymes are described above. The temperature profile may be adjusted depending on the particular template being processed. For chimeragenesis on templates with very long gaps, the sample may need to be held at 60°C for an additional period of time, *e.g.*, 15 to 30 minutes, at the end of the temperature ramp. An alternative method of annealing donor to template is to use single-stranded binding proteins such as recA protein.

Samples or portions of samples may be used to run on denaturing agarose gels to evaluate the extent to which trimming, polymerization, and ligation has restored complete double stranded heteroduplexes. If there is evidence that single-stranded gaps still remain, the TPL steps can be repeated with appropriate adjustments in enzyme concentration, pH, buffer composition, incubation time, or temperature.

At this stage the chimeric heteroduplexes may be transformed into appropriate host cells. Transformation methods are varied and often are specific for the host and/or vector system being used. Methods of transformation are well known to those skilled in the art. Typically, electroporation is used for vectors including BACs. Other methods, including the use of protoplast fusion to DNA:liposome complexes as described above may be used. Transformants are transferred to a rich broth, incubated at the preferred temperature of the host for several hours, and then plated onto the appropriate media for selection. For some applications, it may be advisable to plate transformants onto nonselective media and then to screen (or replate onto selective media) those colonies which arise.

The teachings of all references, patents and patent applications cited herein are hereby incorporated by reference in their entireties. While this invention has been particularly shown and described with references to preferred embodiments thereof, it will be understood by those skilled in the art that various changes in form and details may be made therein without departing from the scope of the invention encompassed by the appended claims.

CLAIMS

What is claimed is:

1. A method for forming a double-stranded chimeric polynucleotide comprising the steps of:

contacting a double-stranded polynucleotide template containing at least two single-stranded regions with a population of oligonucleotides such that at least one oligonucleotide can hybridize to each single-stranded region; and

treating the template and hybridized oligonucleotides, thereby forming a double-stranded chimeric polynucleotide.

2. The method of Claim 1, wherein hybridization is conducted under conditions of low stringency.

3. The method of Claim 1, further comprising a step of filling in gaps.

4. The method of Claim 3, wherein gaps are filled in using a polymerase selected from the group consisting of T4 DNA polymerase, *Taq* DNA polymerase and *Pfu* polymerase.

5. The method of Claim 1, further comprising a step of trimming flaps.

6. The method of Claim 1, further comprising preparing the double-stranded template containing at least two single-stranded regions by modifying at least one strand of the double-stranded template and treating the modified strand such that at least two single-stranded regions are formed.

7. The method of Claim 6, wherein the double-stranded polynucleotide is modified by introducing nicks in a strand of the polynucleotide.

8. The method of Claim 7, wherein the nicked strand is treated with a nuclease to form single-stranded regions.

9. The method of Claim 1, wherein the template and hybridized oligonucleotides are treated with a ligase.

10. The method of Claim 1, further comprising a step of selecting or screening the chimeric polynucleotide for a specified characteristic, wherein said specified characteristic is altered in comparison to a reference polynucleotide.

11. The method of Claim 1, wherein the double-stranded template is at least about 100 kilobases in length.

12. The method of Claim 11, wherein the double-stranded template is from about 100 kilobases to about 350 kilobases in length.

13. The method of Claim 1, wherein the double-stranded template is at least about 10 megabases in length.

14. The method of Claim 1, wherein shearing of the template is minimized.

15. The method of Claim 14, wherein the process is conducted in a semi-solid medium.

16. The method of Claim 15, wherein the semi-solid medium is agarose.

17. The method of Claim 1, wherein the double-stranded polynucleotide template comprises at least one region of random sequence.

18. The method of Claim 1, wherein the double-stranded polynucleotide template comprises more than one gene.

19. The method of Claim 18, wherein the double-stranded polynucleotide template comprises an operon.

20. The method of Claim 1, wherein the double-stranded template comprises genomic DNA.

21. The method of Claim 1, wherein the double-stranded template is circular.

22. The method of Claim 1, wherein the double-stranded polynucleotide template and at least one of the oligonucleotides encodes a desulfurizing enzyme or fragment thereof.

23. The method of Claim 1, wherein a single-stranded region is located in each strand of the double-stranded template.

24. The method of Claim 1, wherein the single-stranded region is at least about 250 nucleotides in length.

25. The method of Claim 1, wherein the single-stranded region is from about 250 to about 5000 nucleotides in length.

26. The method of Claim 1, wherein at least one of the oligonucleotides comprises a region of random sequence.

27. The method of Claim 1, wherein the population of oligonucleotides comprises single-stranded molecules.

28. A chimeric polynucleotide prepared according to the method of Claim 1.

29. A method for forming a double-stranded chimeric polynucleotide comprising the steps of:

preparing a double-stranded polynucleotide template such that at least two single-stranded nicks are introduced;

treating the nicked double-stranded polynucleotide such that a single-stranded region is formed at each nick;

contacting the polynucleotide template comprising the single-stranded regions with a population of oligonucleotides such that at least one oligonucleotide hybridizes to each single-stranded region; and

treating the template and hybridized oligonucleotides, thereby forming a double-stranded chimeric polynucleotide.

30. A method for forming a double-stranded chimeric polynucleotide comprising the steps of:

preparing a double-stranded polynucleotide such that at least one single-stranded nick is introduced;

treating the nicked double-stranded polynucleotide such that a single-stranded region is formed at each nick;

contacting the template comprising the single-stranded region with a population of oligonucleotides such that at least one oligonucleotide hybridizes to the single-stranded region;

trimming flaps of the hybridized oligonucleotides;

filling in gaps of the trimmed, hybridized oligonucleotides; and

treating the trimmed, hybridized oligonucleotides, thereby forming a contiguous chimeric polynucleotide.

31. The method of Claim 30, wherein hybridization is conducted under conditions of low stringency.

32. The method of Claim 30, wherein gaps are filled in using a polymerase selected from the group consisting of T4 DNA polymerase, *Taq* DNA polymerase and *Pfu* polymerase.

33. The method of Claim 30, wherein the single-stranded nick is introduced using DNase I.

34. The method of Claim 30, wherein the nicked double-stranded polynucleotide is treated with a nuclease to form single-stranded regions.

35. The method of Claim 30, wherein the template and hybridized oligonucleotides are treated with a ligase.

36. The method of Claim 30, further comprising a step of selecting or screening the chimeric polynucleotide for a specified characteristic, wherein said specified characteristic is altered in comparison to a reference polynucleotide.

37. The method of Claim 30, wherein the double-stranded template is at least about 100 kilobases in length.

38. The method of Claim 37, wherein the double-stranded template is from about 100 kilobases to about 350 kilobases in length.

39. The method of Claim 30, wherein the double-stranded template is at least about 10 megabases in length.

40. The method of Claim 30, wherein shearing of the template is minimized.

41. The method of Claim 40, wherein the process is conducted in a semi-solid medium.

42. The method of Claim 41, wherein the semi-solid medium is agarose.

43. The method of Claim 30, wherein the double-stranded polynucleotide template comprises at least one region of random sequence.

44. The method of Claim 30, wherein the double-stranded polynucleotide template comprises more than one gene.

45. The method of Claim 43, wherein the double-stranded polynucleotide template comprises an operon.

46. The method of Claim 30, wherein the double-stranded template comprises genomic DNA.

47. The method of Claim 30, wherein the double-stranded template is circular.

48. The method of Claim 30, wherein the double-stranded polynucleotide template and at least one of the oligonucleotides encodes a desulfurizing enzyme or fragment thereof.

49. The method of Claim 30, wherein at least two single-stranded regions are generated.

50. The method of Claim 49, wherein a single-stranded region is located in each strand of the double-stranded template.

51. The method of Claim 30, wherein the single-stranded region is at least about 250 nucleotides in length.

52. The method of Claim 30, wherein the single-stranded region is from about 250 to about 5000 nucleotides in length.

53. The method of Claim 30, wherein at least one of the oligonucleotides comprises a region of random sequence.

54. The method of Claim 30, wherein the population of oligonucleotides comprises single-stranded molecules.

55. A chimeric polynucleotide prepared according to the method of Claim 30.

56. A method for preparing a double-stranded polynucleotide template suitable for use in forming a double-stranded chimeric polynucleotide, comprising the steps of:

treating a double-stranded polynucleotide such that at least two single-stranded nicks are introduced; and

treating the nicked polynucleotide such that a single-stranded region of the polynucleotide is formed at each nick, thereby resulting in a double-stranded polynucleotide suitable for use in forming a chimeric polynucleotide.

57. The method of Claim 56, wherein the double-stranded polynucleotide is obtained by a method selected from the group consisting of: isolating a polynucleotide from a suitable nucleic acid source, synthetically manufacturing a polynucleotide, cleaving the polynucleotide from a larger polynucleotide and amplifying a polynucleotide obtained by any of these methods.

58. The method of Claim 56, wherein at least two single-stranded regions are formed.

59. The method of Claim 58, wherein the single-stranded regions are located on each strand of the double-stranded polynucleotide.

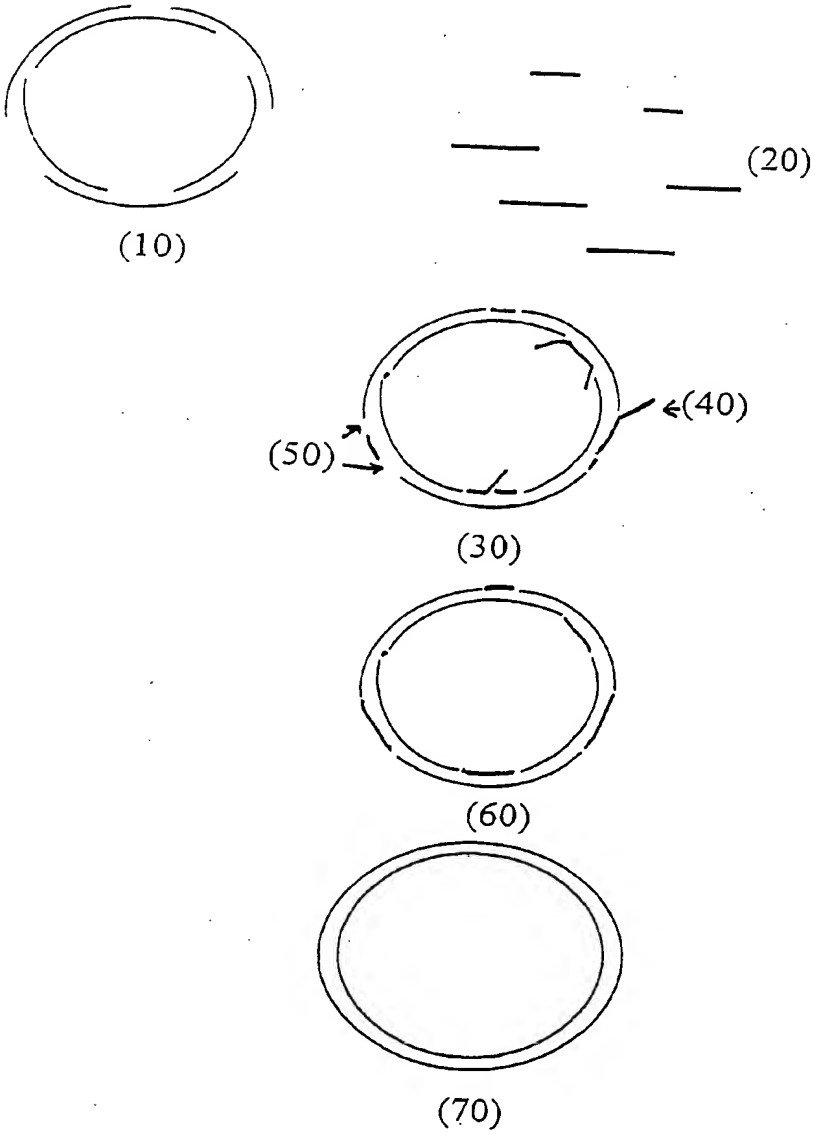


Fig.

INTERNATIONAL SEARCH REPORT

Internatic Application No
PCT/US 00/41319

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/10 C12Q1/68 C12P19/34

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N C12Q C12P

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
WPI Data, PAJ, CAB Data, STRAND, EPO-Internal, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 91 06645 A (MAX PLANCK GESELLSCHAFT) 16 May 1991 (1991-05-16) the whole document ---	30,31, 34-36, 47,51, 54,55
X	WO 98 42832 A (SHAO ZHIXIN ;ZHAO HUIMIN (CN); AFFHOLTER JOSEPH A (US); ARNOLD FRA) 1 October 1998 (1998-10-01) claims 1-34; figure 1 ---	28,55
X	WO 98 27230 A (MAXYGEN INC ;PATTEN PHILLIP A (US); STEMMER WILLEM P C (US)) 25 June 1998 (1998-06-25) page 30, line 15 - line 32; claims 1-274 --- -/--	28,55

[X] Further documents are listed in the continuation of box C. [X] Patent family members are listed in annex.

* Special categories of cited documents :

'A' document defining the general state of the art which is not considered to be of particular relevance
'E' earlier document but published on or after the international filing date
'L' document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
'O' document referring to an oral disclosure, use, exhibition or other means
'P' document published prior to the international filing date but later than the priority date claimed

'T' later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
'X' document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
'Y' document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
'&' document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
12 March 2001	21/03/2001

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Hornig, H

2

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Internatl Application No
PCT/US 00/41319

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
P,X	COCO W M ET AL: "A novel method for shuffling chimeragenesis of mutated genes or of gene families from divergent bacterial genera." ABSTRACTS OF THE GENERAL MEETING OF THE AMERICAN SOCIETY FOR, vol. 100, 2000, page 497 XP000990036 100th General Meeting of the American Society for Microbiology;Los Angeles, California, USA; May 21-25, 2000, 2000 ISSN: 1060-2011 Division O: Abstract no. O-3; abstract	28,55
A	US 4 959 312 A (SIROTKIN KARL M) 25 September 1990 (1990-09-25) the whole document	
A	STEMMER W P C ET AL: "Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides" GENE,NL,ELSEVIER BIOMEDICAL PRESS. AMSTERDAM, vol. 164, no. 1, 1995, pages 49-53, XP004041916 ISSN: 0378-1119 the whole document	
A	WO 93 01282 A (BERLEX LAB) 21 January 1993 (1993-01-21) the whole document	
A	WO 91 06643 A (CANADA MAJESTY IN RIGHT OF) 16 May 1991 (1991-05-16) the whole document	
A	WO 98 28443 A (DADE BEHRING MARBURG GMBH ;ULLMAN EDWIN F (US)) 2 July 1998 (1998-07-02) the whole document	
A	LYAMICHEV V ET AL: "Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes" NATURE BIOTECHNOLOGY,US,NATURE PUBLISHING, vol. 17, no. 3, March 1999 (1999-03), pages 292-296, XP002121762 ISSN: 1087-0156 the whole document	

2

INTERNATIONAL SEARCH REPORT

Information on patent family members

Internati. Application No
PCT/US 00/41319

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9106645 A	16-05-1991	DE 3936258 C	25-04-1991
WO 9842832 A	01-10-1998	US 6153410 A	28-11-2000
		AU 6772598 A	20-10-1998
		AU 724698 B	28-09-2000
		AU 6942098 A	20-10-1998
		BR 9804791 A	17-08-1999
		CN 1230990 T	06-10-1999
		EP 0975653 A	02-02-2000
		EP 0920496 A	09-06-1999
		JP 2000511783 T	12-09-2000
		PL 330287 A	10-05-1999
		WO 9842728 A	01-10-1998
WO 9827230 A	25-06-1998	AU 5729298 A	15-07-1998
		EP 0946755 A	06-10-1999
US 4959312 A	25-09-1990	NONE	
WO 9301282 A	21-01-1993	AU 2317192 A	11-02-1993
		EP 0592597 A	20-04-1994
		JP 6508757 T	06-10-1994
		US 5702931 A	30-12-1997
WO 9106643 A	16-05-1991	US 5071743 A	10-12-1991
		AT 125297 T	15-08-1995
		CA 2071892 A	28-04-1991
		DE 69021073 D	24-08-1995
		EP 0497798 A	12-08-1992
		JP 5502788 T	20-05-1993
WO 9828443 A	02-07-1998	CA 2246225 A	02-07-1998
		EP 0904412 A	31-03-1999